Assignment 03
CSC354 - Semester Project in Scikit-Learn

FA20-BCS-040
FA20-BCS-B
Ahmad Bin Anwar

Loan Prediction System

- **Evaluation Methodology**
  - Run experiments using
    - Train-Test Split Ratio of 80%-20%
    - k-Fold Cross Validation

- **Evaluation Measures**
  - Accuracy
  - Precision
  - Recall
  - F_1

- **Machine Learning Algorithms**
  - Logistic Regression
  - Rain Forest Classifier
  - Naïve Bayes
  - Decision Tree Algorithm
  - Support Vector Machine
  - XG Boost
  - K-Nearest Neighbors (KNN) algorithm
  - Ada Boost Classifier
  - Linear Discriminant Analysis (LDA)
  - Quadratic Discriminant Analysis (QDA)

Tasks

- **Summarize your results** in a Table both for
    - **Train Test Split Ratio**

Machine Learning Algorithm	**F1 Score**	**Accuracy**	**Precision**	**Recall**
Logistic Regression	90%	85%	84%	96%
Rain Forest Classifier	89%	84%	83%	95%
Naïve Bayes	89%	85%	85%	94%
Decision Tree Algorithm	81%	74%	82%	80%
Support Vector Machine	81%	68%	68%	10%
XG Boost	85%	79%	82%	89%
K-Nearest Neighbors (KNN) algorithm	73%	61%	69%	77%
Ada Boost Classifier	86%	80%	83%	90%
Linear Discriminant Analysis (LDA)	90%	85%	83%	99%
Quadratic Discriminant Analysis (QDA)	89%	85%	85%	94%

- **k-Fold Cross validation**

Machine Learning Algorithm	**F1 Score**	**Accuracy**	**Precision**	**Recall**
Logistic Regression	86%	80%	76%	10%
Rain Forest Classifier	90%	82%	83%	93%
Naïve Bayes	90%	85%	85%	95%
Decision Tree Algorithm	85%	80%	90%	81%
Support Vector Machine	83%	70%	70%	10%
XG Boost	87%	80%	83%	91%
K-Nearest Neighbors (KNN) algorithm	78%	66%	71%	86%
Ada Boost Classifier	89%	84%	84%	95%
Linear Discriminant Analysis (LDA)	90%	85%	84%	98%
Quadratic Discriminant Analysis (QDA)	90%	85%	85%	95%

- **Write a report** which describes
  - **Task**
    The task is about Loan Prediction System how a machine learning model can predict whether to give a loan to a person or not based on some parameters. It eases work for banks as it predicts with maximum 90% accuracy .
  - **Dataset**
    The dataset is taken from Kaggle which is already pre-processed but imbalanced.
    https://www.kaggle.com/code/hafidhfikri/loan-approval-prediction/input
  - **Your observations**
    My observations conclude that it performed well on limited amount of data and make most predictions accurately. It can help banks in real world scenarios to make better decisions.